DOCUMENT RESUME

ED 233 010 SP 022 879

AUTHOR Haarer, Barba G.
TITLE Factors Influencing Physical Fitness Evaluation.
PUB DATE Aug 83
NOTE 36p.
PUB TYPE Information Analyses (070) -- Reference Materials - Bibliographies (131)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cardiovascular System; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Individual Differences; *Mild Mental Retardation; Muscular Strength; Physical Education Teachers; *Physical Fitness; Psychomotor Skills; *Student Evaluation; *Test Validity

ABSTRACT

This annotated bibliography focuses on works that examine areas in which the physical educator can improve the administration of physical fitness tests in the elementary and secondary schools. The first part contains annotations that examine modifications of existing components which measure aspects of muscular and cardiovascular endurance. The second part of the study examines methods used in the physical fitness evaluation process: (1) skinfold assessment; (2) test trials; (3) student involvement; (4) scoring; (5) individual differences; and (6) the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) tests. The last part of the study looks at factors influencing evaluation techniques of the educable mentally retarded student in the regular classroom. Some conclusions are drawn from an analysis of the major findings of the annotated works, and recommendations are made for improvements in testing methods. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED233010

FACTORS INFLUENCING PHYSICAL FITNESS EVALUATION

BY

BARBARA G. HAARER

Graduate Student

Study Directed

by

Charles R. DuVall, Ph.D.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barbara G. Haarer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Submitted in partial fulfillment of the requirements
for the course E591---Exit Project
Indiana University at South Bend
August, 1983

SP 022 819

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Factors in Fitness Testing	2
Statement of the Problem	3
Purpose of the Study	3
Organization of the Study	3
Limitations of the Study	4
Glossary	5
ANNOTATIONS	8
Techniques of Performance	8
Muscular Endurance	8
Cardiovascular Endurance	9
Methods of Evaluation	12
Skinfold Assessment	12
Test Trials	13
Student Involvement	14
Scoring	15
Individual Differences	17
AAHPER(D) Tests	18
Suggestions	19
Evaluating the EMR	21
Motivation	21
Suggestions	22
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	24
Summary	24
Conclusions	27
Recommendations	29
BIBLIOGRAPHY	30

INTRODUCTION

Physical fitness is defined as "the ability of the heart, blood vessels, lungs, and muscles to function at optimal efficiency" (Getchell: 8). Optimal efficiency means the most favorable health needed for the enthusiastic and pleasurable participation in daily tasks and recreation activities. Optimal physical fitness makes possible a life style that the unfit cannot enjoy. People who are physically fit look better, feel better, and possess the good health necessary for a happy and full life (Getchell:8).

Physical fitness has received considerable publicity in the past forty years. Interest in physical fitness was greatly increased as a result of World War II and the Korean Conflict. Statistics released by the Selective Service System regarding the high percentage of rejections for military service have stimulated thinking and planning in this area. President Eisenhower began combating the problem by creating the Council on Youth Fitness at the cabinet level to deal with the problem of low fitness among American children. He also formed the Citizen's Advisory Committee on the Fitness of American Youth to examine and explore the facts and then alerted American Youth on what could and should be done (Olree and others:67).

Today, physical fitness continues to be a major concern. The statistics warrant our concern: 1) 40 percent of our school-age population are classified as being obese; 2) half our population is suffering from one or more chronic diseases (3 out of 10 are under 45 years of age); and 3) twenty-seven million Americans have some type of

heart disease (heart disease accounts for 55 percent of all deaths in this country each year) (Cooper:139, Mayhew:40). People of all ages need to be properly assessed as to their level of fitness before a fitness program can be prescribed.

> Although few people today question the beneficial effects of regular physical activity, we still have a lot to learn about what certain activities contribute to physical fitness. There also is a great need for better methods of evaluating the effectiveness of specific physical fitness programs and for better means of measuring the physical proficiency of individuals. (Fleishman:5)
>
> Stan Musial
> Consultant to President
> Kennedy on Physical Fitness

FACTORS IN FITNESS TESTING

Many criticisms have been raised at fitness tests as in the following examples: 1) sit-ups are not an accurate measurement of trunk strength; 2) the influence of body weight on the flexed-arm hang or pull-up; 3) the 600 yard run is too short to measure cardiovascular endurance; 4) unfamiliarity of testees with test items; 5) improper scoring; 6) improper administration of test components; and 7) lack of motivation on the part of the tester or testee (Plowman and Falls:22, Vincent and Britten:74, Baumgartner 1978:80, Meyers:72). Educators are being made aware of these criticisms and need to start acting upon them to obtain accurate assessment of each individual student.

Obtaining an accurate assessment of the performance levels of handicapped students is a major concern to educators who seek to be in compliance with the "Education for All Handicapped Children Act of 1975

(P.L. 94-142). The law states that an individualized education program (IEP) must include a statement of the student's present level of educational performance" (Bowman and Dunn:449). Obtaining this information can be a formidable task. There are many variables that may affect a child's performance at any given moment on any given task.

STATEMENT OF THE PROBLEM

Recently physical fitness tests have been criticized because they tend to disregard external variables which have a definite influence on the outcome of physical fitness testing. Moreover because of the differences of opinion as to what constitutes physical fitness, many fitness tests have been devised. No matter which fitness test is being used, however, violation of sound testing procedures can lead to invalid findings and erroneous interpretation regarding fitness appraisal of the individual.

PURPOSE OF THE STUDY

The purpose of the study is to identify areas in which the physical educator can improve upon when administering physical fitness tests, and thereby to provide information and recommendations to educators who are involved in the fitness testing areas.

ORGANIZATION OF THE STUDY

The first part of the study contains annotations that examine modifications of existing components which measure aspects of muscular and cardiovascular endurance.

The next and major part of the study examines methods used in the evaluation process: skinfold assessment, test trials, student involvement, scoring, individual differences, the American Association of Health, Physical Education and Recreation (Dance) or AAHPER(D) tests, and suggestions for improvement.

The last part of the study looks at factors influencing evaluation techniques of the educable mentally retarded (EMR) student in the regular classroom. Motivation and suggestions for improvement are the key topics examined in this section.

Where annotations are related, they are arranged according to their chronological age.

LIMITATIONS OF THE STUDY

There was not a great deal of information to be found on improving techniques of performance or incorporating the educable mentally retarded into the regular classroom for fitness testing. Also, there were only a few issues where controversy was presented with research being done on opposing viewpoints. It was not the intent of this study to examine all of the many components that could go into a fitness test. Likewise, it was not the intent to research all of the fitness tests that have been written. The only tests given much mention were the AAHPER(D) tests because of their widespread use.

GLOSSARY

AAHPER(D) -- American Association of Health, Physical Education, and Recreation (Dance).

aerobic -- "living, occuring, or acting only in the presence of oxygen," (Webster:19). This refers to an exercise performed for a long period of time.

aerobic capacity -- "the largest amount of oxygen that can be consumed per minute," (Getchell:24). Synonym: work capacity.

agility -- "[illegible] ability to make a sudden change of direction or movement," [illegible]asicker:355).

anaerobic -- "[illegible]iving, occurring, or acting in the absence of free oxygen," (Webster:41). This refers to an exercise performed for a short period of time.

cardiorespiratory -- "of or relating to the heart and the respiratory system," (Webster:168).

cardiovascular -- "of, relating to or involving the heart and blood vessels," (Webster:168).

cardiovascular endurance -- "the capacity of the individual to maintain strenuous activity of a number of muscle groups or of the whole body for a prolonged period." This could be subdivided into aerobic and anaerobic capacity. (Simri:374).

correlation -- "refers to the relationship of one variable to another," (Johnson and Nelson:25).

correlation coefficient -- "a measure of the relationship of two random variables that range between +1 and -1, or from a perfect positive relationship tc a perfect negative relationship," (Johnson and Nelson:25).

criterion -- "a standard on which a judgment or decision may be based," (Webster:270).

dynamic strength -- "ability of the individual to move or support the weight of the body," (Simri:364).

Educable Mentally Retarded (EMR) -- "children with below normal intelligence; usually possessing an I.Q. from 55-69," (Lowenbraun and Affleck:9).

endurance -- "the ability to apply strength and sustain it," (Getchell:10).

ergometer -- "an apparatus for measuring the work performed by a group of muscles," (Webster:388).

explosive strength -- "the ability to release maximum force in the fastest possible time," (Simri:367).

flexibility -- "the ability to use a muscle throughout its maximum range of motion," (Getchell:10).

mean -- "the arithmetic average," (Johnson and Nelson:19).

muscular endurance -- "the ability of the individual to move or support the weight of the body repeatedly and continuously," (Simri:364).

musculoskeletal -- "of, relating to, or involving both musculature and skeleton," (Webster:758).

norms -- "values considered to be representative of a specified population," (Johnson and Nelson:44).

psychomotor -- "of or relating to motor action directly proceeding from mental activity," (Webster:931).

reliability -- "accuracy and repeatability of a test," (Nelson and Dorociak:63).

static strength -- "the exertion of a maximum force for a brief period of time against a fairly immovable object," (Simri:369).

strength -- "the capacity of a muscle to exert a maximal force against a resistance," (Getchell:9).

validity -- "the test measures what it is intended to measure," (Nelson and Dorociak:63).

ANNOTATIONS

TECHNIQUES OF PERFORMANCE

Muscular Endurance

Vincent, W. J., and Britten, S. D., "Evaluation of the Curl-Up: A Substitute for the Bent Knee Sit-Up," Journal of Physical Education and Recreation 51:74-75, February, 1980.

The authors of this article agreed upon and suggested that the bent knee sit-up is unacceptable as an exercise for at least three reasons. First, during the sit-up the abdominal muscles account for only the first portion of the movement. The remaining movement is accomplished by the hip flexors. Second, the strengthening of the hip flexors is usually not necessary. Third, the additional strengthening of the hip flexors is harmful because these muscles are required to contract while in an already shortened position. For these reasons the authors attempted to devise a substitute for the sit-up. This resulted in what is called the "curl-up." The subject assumes the supine lying position with knees bent and hands behind the head. The abdominals are then contracted to raise the upper back off the mat just slightly. The authors conducted tests in school settings to test the mass testability and reliability of the curl-up. They concluded that the curl-up was not a reliable measure of abdominal strength. However, the curl-up appeared to be a better exercise for daily abdominal muscle development. The sit-up remains as the only available instrument to measure abdominal strength.

Baumgartner, T. A., "Modified Pull-Up Test," Research Quarterly 49:80-84, March, 1978.

A modified pull-up test was devised by this author to substitute for the flexed arm hang test which is the arm and shoulder strength test used in most physical fitness tests. The problem with this flexed arm hang test, according to the author, is that light rather than heavy subjects seem to do well on this test since body weight is the resistance. An added problem to this test is that many subjects cannot do even one pull-up. The modified test uses a pull-up bar constructed on one end of a board. This end of the board is placed on a doorway pull-up bar positioned five feet above the floor. The other end of the board is placed on the floor so that the board is at a 30 degree angle to the floor. As the subject lays forward onto the inclined board, the test administrator places a four-wheel scooter board under the subject's midsection. The test administrator and an assistant push the subject on the scooter board up to the top of the inclined board. The subject grasps the pull-up bar with an overhanded grip with hands held shoulder-width apart. The subject pulls up the inclined board until the chin is over the pull-up bar and returns to the straight-arm hanging position. The subject repeats this as many times as possible. The author concluded by saying that he has found this test to be quite valid and reliable as it discriminates among ability levels.

Cardiovascular Endurance

Adams, F. A.; Linde, L. M.; and Miyake, H., "The Physical Working Capacity of Normal School Children in California," Pediatrics 28:55-64, 1961.

The authors of this article gave a bicycle ergometer test to 243 normal white school children of Southern California of both sexes ranging in age from 6 through 14. Each student's surface area was calculated using the formula "surface area = $\text{weight}^{0.425} \times \text{height}^{0.725} \times 71.84$. In the test each student was asked to perform consecutively three different work-load trials on an electric bicycle ergometer. The pedaling rate was maintained between 60 and 70 rev/min., and each workload trial lasted 6 minutes. An attempt was made to schedule the work loads of each subject so that the first produced a heart rate from 100-120 beats per minute, the second produced 120-140 beats per minute, and the third 150-170 beats per minute. The authors found that boys possessed significantly greater working capacities than girls even at the smaller surface area. Differences became quite marked in the older and larger boys as contrasted with girls of the same age or size. They feel that this working-capacity test has a number of advantages over other methods of evaluating cardiovascular fitness. It is objective, reproducible, and easily administered.

Cotton, Doyice J., "A Modified Step Test for Group Cardiovascular Testing," Research Quarterly 42:91-95, March, 1971.

The purpose of this study was to develop a modification of the Ohio State University Step Test of cardiovascular fitness which would be satisfactory for mass testing in a classroom. The author came up with a modified step test which was administered to 46 boys in grades 9-12. The test consists of 18 consecutive innings of stepping on a 17" bleacher step. Each inning consists of 30 seconds of work and 20 seconds of rest with no pause in between innings (total 15 minutes). During the 20 second rest period the subject's pulse is taken at the carotid artery for 10 seconds beginning with second "5" and ending at second "15." A buddy system in which one partner exercises and the other counts the pulse is used so that half of the class can be tested simultaneously. The individual completes the test and ceases to exercise when the subject's heart rate reaches 150 beats per minute (BPM), or when the subject completes 18 innings. The score is the inning in which the heart rate reaches 150 BPM. If 18 innings are completed, the subject is assigned a score of 19. The test-retest reliability is quite high; therefore, Cotton concludes that it is a valid and easy-to-administer test for cardiovascular endurance. However, the author feels that modifications need to be made for female subjects.

Long, Boyde, "To Test Endurance," Athletic Journal 53:80-82, March, 1973.

This physical educator from Bellingham, Washington, felt that the twelve minute run-walk is a good test for speed, but that it does not accurately measure endurance. The test that he used was a 40 yard dash test in which each student was required to run ten 40 yard dashes as fast as he/she possible could. The student was given 15 seconds of rest between dashes. In evaluation of this test the instructor was required to take the first two 40 yard dash times and find the average for the target time. Next the instructor added the times of the remaining eight dashes and divided by eight. This total was then divided into the figure attained for the target time to get a percentage. The object was to keep the percentage as close to 1.00 as possible if the student had a reasonable amount of endurance. The author concluded that this test is good to test endurance as the heavier student can score as well as the leaner student, and the test does not discriminate among the speed abilities of the various participants.

Cooper, Kenneth H., "Testing and Developing Cardiovascular Fitness," Journal of Physical Education Special Edition:130-144, March-April, 1972.

Dr. Ken Cooper, of the Institute for Aerobics Research of Dallas, completed studies that have shown that in runs completed in less than 8 to 10 minutes much of the energy used comes from anaerobic energy sources. He found that after a "best-effort" one minute run, slightly more than 60 percent of the work is performed anaerobically. At 10 minutes of work, less than 10 percent of the work is anaerobic. He stated that if maximum work is performed for 10 to 20 minutes, the predominant energy source is dependent on the utilization of oxygen. An adequate estimate of one's aerobic capacity will, therefore result. Consequently, he concluded by saying the 12 minute or 1½ to 2 mile run has proven to be an effective indicator of cardiovascular fitness.

Disch, James; Frankiewicz, Ronald; and Jackson, Andrew, "Constant Validation of Distance Run Tests," Research Quarterly 46:169-176, May, 1975.

This study investigated the construct validity of distance run tests. A sample of 60 male college students were administered 10 running tests which ranged in distance from a 50 yard dash to the 2 mile and 12 minute run. The authors wanted to determine if a run of a specified distance measured primarily speed or some type of endurance. They found that the correlation coefficients between aerobic capacity and distances of 440 yards and shorter were not significantly different from zero. In contrast, the correlations with distance runs of 600 yards and longer tended to be significantly from zero. The authors concluded by saying that the 2 mile and 12 minute runs were the most valid in measuring aerobic capacity.

METHODS OF EVALUATION

Skinfold Assessment

Mayhew, Jerry L., "Body Composition," Journal of Physical Education, Recreation, and Dance 52:38-40, September, 1981.

The advent of the new AAHPER(D) Fitness Test has brought increased interest in body composition measurement. Body composition refers to the types and amounts of tissues which make up the body. The most widely accepted model is the two-component scale which divides the body weight (BW) into lean body weight (LBW) and fat weight (FW). The relative fat content, or percent of fat, is simply the ratio of FW to BW expressed as a percent. This author stated that skinfold measurement using skinfold calipers is a simple and accurate way to measure an individual's percentage of fat in a classroom setting. He stated that once the skinfold has been secured between the thumb and forefinger, simply apply the jaws of the caliper to the middle of the pinch. Pressure should then be released on the caliper allowing the jaws to settle on the site of evaluation. This should be repeated several times to obtain the mean value for each site. Each student can then estimate his/her ideal body weight by dividing his/her lean body weight by 1 (optimal % fat/100). The optimal percentage of fat for males is 10-14% and 20-24% for females. The author concluded by saying that with as much as 40% of the school-age population being obese, it is vital to the health of the nation that each student be aware of his/her body composition.
[Mayhew is the director of the Human Performance Laboratory at Missouri State University.]

Lohman, Timothy G., and Pollock, Michael L., "Which Caliper? How Much Training?" Journal of Physical Education and Recreation 52:27-29, January, 1981.

The authors of this article conducted a study to test three different types of skinfold calipers between experienced and inexperienced testers. Fifteen children between 8 and 14 years of age were measured at the locations of the upper arm (triceps) and back (subscapular). The three caliper types were: 1) a plastic caliper with neither spring nor tension; 2) a plastic caliper with a spring; and 3) a metal capiler with uniform tension. Calipers 1 and 2 produced low correlations between the experienced and inexperienced tester for the triceps and a low correlation for caliper 1 for the subscapular area. For caliper 3 the results were somewhat better, indicating the better designed caliper can be used more successfuly with inexperienced testers. After training the inexperienced testers, the measurements were repeated one week later. At this time all three calipers proved to be satisfactory based on the correlations. This study demonstrated that when inexpensive calipers are used by inexperienced testers, they might not yield accurate skinfold measurements.

hawkins, Jerald D., "An Analysis of Selected Skinfold Measuring Instruments," Journal of Physical Education, Recreation, and Dance 54:25-27, January, 1983.

Hawkins did a study to compare the reliability of three newer and less expensive skinfold calipers to the more expensive Lange Skinfold Caliper. The three less expensive calipers were the Slim-Guide, Fat-O-Meter, and Adipometer. The subjects of this study were 800 volunteers ranging in age from elementary up to university students. The subjects in each age group were randomly assigned to one of two test groups for the purpose of conducting two separate testing sessions. The triceps area was chosen as the area for skinfold measurement and each subject was tested with all four instruments. All measurements were conducted by a single experienced tester who systematically varied the order of instrument use. The results failed to indicate significant differences among the Lange, Fat-O-Meter, and Adipometer. However, the Slim-Guide instrument consistently produced measurement values significantly lower than those of the other three models. Hawkins concluded by saying that the Fat-O-Meter and Adipometer skinfold calipers should be acceptable alternatives to the more expensive Lange caliper, especially when used by an experienced tester in a class testing environment.

Test Trials

Liba, Marie, "A Trend Test as a Preliminary to Reliability Estimation," Research Quarterly 33:245-248, 1962.

The purpose of this study was to determine whether trial-to-trial variation was systematic or random in two sets of performance data. Two sets of performance data--velocity of the ball in bowling and velocity of the overhand throw--for two different groups of college women were obtained. All subjects were members of regular classes in these activities, and sets of performance data were collected near the end of an eight-week unit of instruction. Five trials recorded on a single day were used for this study. Subjects in the bowling group had been tested regularly, and subjects in the softball group had considerable practice in throwing for distance. Two warm-up trials preceded the five test trials. The author found that there was systematic variation in the bowling data but only random variation in the softball data. The author concluded by saying that when testing any item the first step should be to administer a trend test of this type to determine reliability. When there is evidence of systematic variation from trial to trial, the model is not appropriate.

Baumgartner, Ted A., and Jackson, Andrew S., "Measurement Schedules for Tests of Motor Performance," Research Quarterly 41:10-14, 1970.

These authors tested five groups of male subjects on a variety of physical performance tests. They wanted to determine the number of test trials that were necessary to maximize group performance. They found that when administering jumping and agility tests to junior high boys, two or three trials were sufficient due to the fatigue factor. For the high school boys and college men, five or six trials needed to be administered since the later trials resulted in better measurements. For the ball velocity tests, four to five test trials were needed for all age groups to get maximum performance, and two trials were adequate for all age groups in the 50 yard sprint due to fatigue. In testing for static muscular strength, three trials were sufficient for arm strength. However, four or five trials had to be administered for the leg strength tests. The authors concluded by saying that when all trials are to be administered on the same day, more test trials than are suggested in test and measurement books are necessary to achieve a reliable criterion score.

Student Involvement

Martens, Rainer, and Landers, Daniel," Coaction Effects on a Muscular Endurance Task," Research Quarterly 40:733-737, 1969.

These authors wanted to determine the effects of one and three coactors on an individual's performance of a muscular endurance task. The subjects used in this study were 8-, 13-, and 18-year old males. For this test, subjects extended the dominant leg in a horizontal position for as long as possible under one of three conditions: alone, in pairs, and in groups of four. Results showed that individuals in quadrads performed significantly better than individuals performing alone or in dyads. The authors suggested that perhaps the presence of some minimum number of individuals was needed to produce sufficient arousal to significantly alter performance. On the basis of this study, it could only be assumed that one coactor is not sufficiently arousing, but three coactors are. The authors also suggested that individuals begin to compete with members of their group even when they are told not to compete. Also, it might be possible that increased arousal resulting from the presence of other individuals expanded a person's physiological limit. In conclusion, they stated that physical educators must not only be cognizant of physiological and kinesiological factors in motor performance but also social factors.

Nelson, J. K., and Dorociak, J. J., "Reducing Administration Time While Improving Reliability and Validity of Fitness Tests," Journal of Physical Education, Recreation, and Dance 53:63-65, January, 1982.

In the introductory jogging classes at Louisiana State University, students were tested on the following: 1½ mile run, 1 minute sit-ups, and sit and reach flexibility as part of the evaluation process. A study was done to test the reliability of each test item by using a rehearsal group compared to a traditional group in two test trials. The traditional group was given no formal practice of the test items. The rehearsal group was given several minitests of each test item and a couple of trials with the actual test before the first testing day. The rehearsal group also utilized partners to help score the test and to assist in pacing the practice sessions. The authors found that the rehearsal group subjects exhibited more reliable performance from test to retest than the traditional group. Also, there were greater changes from the first to second test administrations for the traditional group. They feel that the involvement of the students in the testing process has value in terms of validity, reliability, administrative efficiency, and as an educational experience for the students.

Scoring

Hetherington, Ross, "Within-Subject Variation, Measurement, Error, and Selection of a Criterion Score," Research Quarterly 44:113-117, 1973.

This author questioned the use of the mean to estimate maximum performance. It seemed to Hetherington that the experimenter should use average scores when the major source of unreliability is measurement error and use maximum scores when within-subject variation is high as compared to measurement error. He did a study with 105 fifth grade boys to test his hypothesis. They were tested for ability to exert maximum elbow flexion muscular force with the elbow at 60, 90, and 120 degrees. The example presented in this study demonstrated that the estimate of within-subject variation was large as compared to the estimate of measurement error. Therefore, the boys in the study had a better score when the "best" score was determined as to when the mean score was determined. Hetherington concluded by saying that in the measurement of maximum physical performance, when it is believed that measurement error is small relative to within-subject variation, the best score obtained on any trial represents the best prediction of the true score.

Baumgartner, Ted A., "Criterion Score for Multiple Trial Measures," Research Quarterly 45:193-197, 1974.

This article gave a brief review of various articles dealing with whether to use the best score or average score as the criterion score for multiple trial physical performance tests. Baumgartner wanted to determine whether multiple trial scores for a subject were normally distributed. Approximately 260 male subjects were administered four multiple trial tests on each of two days. Based on the information gathered, the author stated that it seems impossible to say that the use of one method results in higher correlations than the other. Only in two comparisons (junior high reaction time and college reaction time) were the differences between the two correlation coefficients large enough to have practical applications. Baumgartner further stated that the following conclusions seem warranted: 1) the multiple trial data of a subject taking a physical performance test are usually normally distributed; and 2) in almost all situations the use of a criterion score based on the mean of the trial scores is more defensible than the use of the best score.

Cooter, Rankin G., "Who Wants to be Normal?" Journal of Physical Education and Recreation 47:50, May, 1976.

This author stated that instead of setting excellence as the goal for students, physical educators too often tend to refer to norms as if they were the ideal for such characteristics as weight and physical fitness. Cooter feels that student scores on the AAHPER(D) Youth Fitness Test or other motor fitness tests are being compared to norm values and misused for grading purposes. He stated that if the physical educator is not extremely careful in interpreting the norms, students may be unduly penalized or become discouraged by the results. He also stated that the instructor should use these norms only for motivation and as general guidelines to the level of performance the youngster has achieved. Physical educators, too, need to strive for local, representative norms and be constantly updating them. In conclusion, the author stated that physical educators need to get away from grading and comparing students on a national norm if they want to create an atmosphere in which desirable levels of health and physical well being are sought. [Cooter is director of the Human Performance Laboratory at Georgia State University.]

Londeree, Ben R., "Strength Testing," Journal of Physical Education, Recreation, and Health 52:44-46, September, 1981.

This author stated that physical educators need to look at students' scores on the AAHPER(D) Youth Fitness Test and prescribe exercises to improve problem areas to students who fall below the fortieth percentile on any of the test items. Londeree stated that many postural deviations result from strength and/or flexibility imbalances, and that physical educators are in a position to use this knowledge from fitness testing to correct these conditions. Also, each student should have an evaluation taken of his/her posture from a side view. If body parts do not line up, proper exercises should be prescribed. Students who score below the fortieth percentile on the leg jump test should have knee extension and flexion exercises prescribed. Students who fall below the norms on flexed arm hang and knee push-ups should work on these individual exercises. However, Londeree feels that educators need to consider body composition and prescribe weight loss in addition to these exercises. Londeree concluded by saying that the high injury rate experienced by exercisers necessitates an active preventative approach rather than ignoring the problem.
[Londeree is the director of the Human Performance Laboratory at the University of Missouri.]

Individual Differences

Baumgartner, Ted, and Zuidema, Marvin A., "Factor Analysis of Physical Fitness Tests," Research Quarterly 43:443-450, December, 1972.

For the purposes of this study, four factors were hypothesized as constituting the domain of physical fitness: 1) upper body strength and endurance; 2) trunk strength and endurance; 3) leg strength and endurance; and 4) cardiovascular endurance. This study was administered to college men and women to determine if these hypothesized four factors of physical fitness could be supported and whether the factor structure was the same for men and women. It was found that chin-ups most accurately measure upper body strength in men. The half hold sit-up was discovered to be a better measure of trunk strength than the conventional sit-up. Standing broad jump and the 75 yard dash were both equally good in measuring leg strength in the men, and the 12 minute and 880 yard runs were both valid in measuring cardiovascular endurance. For the females the modified pull-up most reliably measured upper body strength, and sit-ups and half hold sit-ups best identified the trunk strength factor. The standing broad jump was most accurate in measuring leg strength, and the Harvard step test best identified cardiovascular endurance. These authors concluded by saying that all four factors were valid in assessing physical fitness for both men and women. However, the components that made up the factors varied as to their reliability in assessment.

Montoye, H. J.; Frantz, M. E.; and Kozar, A.J., "The Value of Age, Height, and Weight in Establishing Standards of Fitness for Children," Journal of Sports Medicine and Physical Fitness 12: 174-179, 1972.

This article stated it is common practice in the schools to use combinations of age, height, weight, and grade to establish standards of performance for physical education activities. In both national surveys of youth fitness in the United States, percentile standards were reported by age and by a classification system based on age, height, and weight. The superiority of the latter method was implied. The authors administered the youth Fitness Test battery and two other fitness tests to virtually all the children in grades 4 through 12 in Tecumseh, Michigan. They wanted to determine if age, height, and weight were of sufficient value to merit the additional time and effort in testing. From the results, the authors concluded that one can disregard height and/or weight, if achievement standards are established for specific ages and for boys and girls separately.

AAHPER(D) Tests

Olree, H., and others, "Evaluation of the AAHPER Youth Fitness Test," Journal of Sports Medicine and Physical Fitness 5:67-71, 1965.

The seven items of the AAHPER Youth Fitness Test (pull-ups, sit-ups, shuttle run, standing broad jump, 50 yard dash, softball throw, and 600 yard run-walk) were administered to 76 males, aged 16-17, according to the directions given in the test manual. Simple correlations were computed for all combinations of the seven items and the correlation coefficients between pull-ups, sit-ups, and the 50 yard dash were very low, indicating that three test items apparently measured different aspects of physical fitness. Each of these three test items was relatively easy to administer. The authors concluded by saying that all three test items may be given in a short period of time with little strain on the individual and will eliminate one of the two days previously required for giving all seven items.

Falls, A. B., "Modern Concepts of Physical Fitness; AAHPERD Health-Related Physical Fitness Test," Journal of Physical Education and Recreation 51:25-27, April, 1980.

During its November 1979 meeting, the AAHPERD Board of Governors gave final approval to the new AAHPERD Health-Related Physical Fitness Test. This test concentrates on health-related fitness which is believed to offer the individual some protection against degenerative type diseases such as coronary heart disease, obesity, and various musculoskeletal disorders. The old test concentrated on performance-related fitness, which included those qualities of function that provided the individual with the ability to participate in sports activities with greater power, strength, endurance, skill, etc. The author feels that the basic emphasis should be placed on health-related fitness or cardiovascular function, body composition, strength, and flexibility. These are also among the basic components in the performance-related fitness area. However, the major differences lie in the degree to which each component must be developed within the two areas. The basic health-related fitness components are measured in the new AAHPERD test by the one mile or nine minute run, sum of subscapular and triceps skinfolds, one minute timed sit-up, and sit and reach, respectively.
[Falls is chairman of the AAHPERD Test Force on Youth Fitness.]

Suggestions

Franklin, Barry, "It's Almost Time for Another Invalid Appraisal of our Nation's Youth Fitness," Physical Educator 29:156, October, 1972.

This author feels that too many physical educators are selecting only one test to use as the determinant for a specific constituent of fitness. The physical educators' rationale behind this single test has many factors: 1) it is easy to score; 2) it is simple to administer; and 3) there is a high correlation between performance on the test and other tests measuring the same component. However, Franklin feels that by selecting this single test criterion as the determinant of a student's level of fitness, the fitness appraiser must also unjustifiably accept the inferred correlation between failure on this test and failure on subsequent other tests also measuring the component in question. He feels that physical educators' goals must therefore lie in developing new and extensive unbiased test batteries consisting of many items per component. Furthermore, these should be individualized by body weight or soma type classification. Franklin stated that only in this manner will we have a true, valid appraisal of our nation's youth fitness.
[Mr. Franklin is a staff member at Pennsylvania State College.]

Simri, Uriel, "Assessment Procedures for Human Performance," in Fitness, Health, and Work Capacity: International Standards for Assessment, pp. 362-377, edited by Leonard A. Larson, MacMillan Publishing Company, New York, 1974.

Simri stated that the order of testing is of utmost importance. He said that speed is best tested at the beginning of the class period and should be followed by the test for explosive strength such as the long jump. The endurance test is usually so demanding that a prolonged rest is recommended following it. Therefore, he stated that the endurance test should be placed at the end of the testing session. He also recommended that scores on the various physical fitness tests should be summarized by establishing one total physical index. He felt that this is unattainable, however, because not all components of the physical fitness test carry the same weight.
[Simri is a director at Wingate Institute of Physical Education and Sports in Netanya, Israel.]

Meyers, C. R., "In Quest of Sound Fitness Testing," Physical Educator 37: 69-75, May, 1980.

The most prevalent shortcomings in fitness testing listed by Meyers were: 1) either lack of or insufficient test orientation to provide full comprehension of the task so that the performer can perform to maximum of ability; 2) inadequate preparation of testees to meet the physical demands of the test task without experiencing detrimental effects during or after performance; 3) lack of willingness in testees to exert their best effort in each test task (Factors influencing willingness include motivational approach, overt behavior of testers, and utilization of a competitive atmosphere.); 4) inaccurate scoring by not affording practice, supervision of scoring, and familiarity with scoring procedure; and 5) not allowing adequate and equitable rest between test items and not following the same order of test items for everyone. Meyers concluded by saying that all such testing should be conducted with due attention to these considerations, because they are essential to give meaningful results.
[Dr. Meyers is professor of Physical Education and director of the Measurement Evaluation Laboratory at the State University of New York.]

EVALUATING THE EMR

Motivation

Wagner, Harry Douglas, The Effects of Motivation and Repeated Trials on Physical Proficiency Test Performance of Educable Mentally Retarded Girls, Unpublished doctor's thesis, George Peabody College for Teachers, August, 1976, 91 pp., typed.

Wagner compared nine educable mentally retarded girls to twelve intellectually normal girls of comparable mental age and to twelve intellectually normal girls of comparable chronological age. He tested them on measures of grip strength, volleyball throw, standing broad jump, and the 30 yard dash. Each was presented under standard instruction, active encouragement, and candy reward motivating conditions. He found that all three groups performed significantly better when active encouragement was added to standard instruction. However, the most dramatic difference was noted when the candy reward was presented to the EMR girls. He also found that the EMR group needed many more trials than the other groups to reach maximum performance. Both normal groups showed a rapid decrease in scoring after reaching maximum performance while the decline for the EMR group was much more gradual. Wagner concluded by saying that the need for motivation and practice is of even greater significance to EMR children than to normal peers.

Levy, Joseph, "Social Reinforcement and Knowledge of Results as Determinents of Motor Performance Among EMR Children," American Journal of Mental Deficiency 78:752-758, 1974.

This study was designed to determine the effect of social reinforcement and knowledge of results on the performance of an accuracy motor task. Subjects were 80 EMR children between the ages of 8-14. The motor task used was performance on a Lafayette Photoelectric Rotary Pursuit. Subjects received 30 trials broken up into six blocks of trials. Each trial lasted 20 seconds followed by a 15 second rest. There was a two minute rest between each of the six blocks of five trials. The trials were broken up and evaluated under the following four conditions: 1) children were given positive social reinforcement; 2) they were given tangible rewards (candy); 3) negative reinforcement was used; and 4) no feedback was given. In all of the four conditions they were given knowledge of their results half of the time and no knowledge of their results for the other half. Levy found that the EMR children scored higher when given a tangible reward with knowledge of results as compared to positive reinforcement with knowledge of results. In all cases the children scored higher when knowledge of results was given as compared to when it was not given. In conclusion, Levy stated that using tangible rewards with knowledge of results will bring out the best performance of an EMR child in any testing situation.

Bowman, R. A., and Dunn, J. M., "Effects of Peer Presence on Psychomotor Measures with EMR Children," Exceptional Children 48:449-451, Fall, 1982.

The purpose of this study was to analyze the effect of the presence of others on three psychomotor measures with EMR children. The three fitness test items were sit-ups, shuttle run, and standing long jump. Subjects were tested twice on the three items. In one test session the experimenter and the student were the only individuals present. In another test session the experimenter and the student performed before a passive audience composed of four of the peers and one adult observer. It was found that the standing long jump was not significantly influenced by the presence of a group. The shuttle run and sit-ups had significantly higher scores when performed before a passive audience. These authors stated that the presence of a group appears to motivate subjects when they are involved in simple motor responses. Consequently, the educator must consider the following major variables when working with educable mentally retarded children: 1) whether the test necessitates a simple or complex motor response; and 2) whether the student could be motivated to perform better if a group were present.

Suggestions

Rarik, Lawrence G.; Widdop, James H., and Broadhead, Geofrey D., "The Physical Fitness and Motor Performance of Educable Retarded Children," Exceptional Children 36:509-519, 1970.

A modification of the AAHPER Youth Physical Fitness Test was administered to a national sample of 4,235 educable mentally retarded boys and girls, ages 8-18. The flexed arm hang was substituted for the pull-ups for boys since many could not perform a single pull-up. The sit-up test was altered from the number of sit-ups executed in an unlimited time to the number which could be executed in one minute. The 300 yard run was substituted for the 600 yard run-walk since the pilot study indicated that the shorter distance was long enough to tax the endurance of the retarded children. A slight change was made in the distance throw in that the measurement was taken from the subject's position at ball release. Age changes in performance followed essentially the same trends as in normal children, although the EMR children were substantially behind standards of performance of normal children. Sex differences in performance of the retarded were similar to those noted in normal children, with boys showing superiority in all tests at all ages. The authors concluded by saying that performance standards for normal children are clearly not appropriate for use with the educable retarded. They stated that norms have to be set up for varying heights, weights, and ages for the seven performance tests.

Jansma, Paul, "Psychomotor Domain Tests for the Severely Handicapped," Journal of the Association for the Severely Handicapped 5:368-381, Winter, 1980.

This author suggested the following guidelines for the appropriate selection and use of tests with the seriously handicapped. These are: 1) data should be gathered on a continuous basis; 2) a definite diagnosis should not be made on the basis of administering a single fitness test; 3) a tester should read the test and know it before using it; 4) the tester should know or be familiar with the child to be tested before the test is administered; 5) the test must test what it proposes to test; 6) the test should be reliable and objective; 7) no progression or scale fits all individuals at each level; 8) the tester should not use the same tests at all times for all students; 9) the gathering of observational data over time is advised; 10) attention to a tested student's chronological age level of performance should decrease in importance as a function of handicap severity; 11) the more seriously a student is handicapped, the more attention that must be paid to extraneous environmental variables; and 12) parents and direct care staff are excellent sources for test data. He concluded by saying that instruction must be preceded by an accurate determination of current developmental level information since the "foundation of instruction is testing and assessment."

SUMMARY

In modifying techniques of performance, Vincent and Britten suggested that the sit-up was not an acceptable exercise, but found that the curl-up was not a reliable substitute. Baumgartner stated that the flexed arm hang was not a reliable exercise for arm and shoulder strength, so he devised a modified pull-up test which does not discriminate among varying weights. Cooper, Disch, Frankiewicz, and Jackson agreed that when measuring cardiovascular endurance the 12 minute and 2 mile runs are most valid. Cotton, Long , Adams, Linde, and Miyake devised substitutes for these long distance runs. Cotton invented a modified step test, Long used ten 40 yard dashes, and Adams, Linde, and Miyake discovered a bicycle ergometer test to measure cardiovascular endurance.

Hawkins, Lohman, and Pollock found through their experiments that the less expensive skinfold calipers could be used successfully by experienced testers. However, Lohman and Pollock discovered a problem when inexpensive calipers are used by inexperienced testers. They found that, in this instance, the results might not yield accurate skinfold measurements. Mayhew stated that it is important for educators to know how to calculate ideal body weight, also.

Jackson and Baumgartner found that evaluation techniques could also be improved by offering more trials than those listed in test manuals. However, they found that junior high subjects could not be administered as many trials in jumping and agility tests due to the fatigue factor. Liba agreed with Jackson and Baumgartner and also stated that many trials are needed to determine if there is systematic variation from trial to

trial. She feels that when there is a lot of variation, the test is not appropriate.

Martens and Landers found that students will perform better in the presence of other students than when performing alone. The presence of three peers had more effect than the presence of a single individual. Nelson and Dorociak also stated that students were an asset in the evaluation process as they could assist each other in practicing for the test items and could also assist in scoring.

In scoring tests, Baumgartner's research found that the use of mean scores is more defensible than the use of the best score except when evaluating reaction time. Hetherington went on to say that when measurement error is small relative to within-subject variation, the best score is the better prediction of the true score. Cooter stated that norms should not be used for comparing students on a national level. Rather, norms should be set up on a local, representative basis. Londeree said that national norms should be used to improve problem areas to students who fall below the fortieth percentile on any of the test items.

When examining individual differences, Baumgartner, Zuidema, Adams, Linde, and Miyake discovered in their studies that the components that make up a physical fitness test should vary as to whether the subjects are male or female. Adams, Linde, and Miyake found females to have smaller working capacities. Contrary to popular belief, Montoye, Frantz, and Kozar found that testers can disregard height and weight in evaluating fitness. However, achievement standards need to be established separately for boys and girls.

The old and new AAHPER(D) fitness tests were mentioned when trying to improve test selection. Olree and others found that in the old AAHPER(D) Fitness Test only pull-ups, sit-ups, and the 50 yard dash measured different aspects of physical fitness. Falls stated that the new AAHPER(D) Fitness Test measures only the one mile or nine minute run, subscapular and triceps skinfolds, sit-ups, and sit and reach. Various criticisms were given by Franklin, Meyers, and Simri that hinder test administration such as: 1) only one test item is being selected for each component of fitness; 2) lack of test orientation; 3) inadequate preparation of testees; 4) lack of motivation in testees; 5) inaccurate scoring; 6)improper order of test items; and 7) not allowing enough rest between items.

In testing the EMR child, both Wagner and Levy found in their studies that the EMR child showed more improvement when provided with motivation in the form of a tangible reward than did the other students. Wagner also discovered that the EMR child needed many more trials to achieve maximum performance. Bowman and Dunn found that the EMR child performed better in simple motor responses, such as sit-ups or a dash, when performing before an audience. Complex motor activities, such as the long jump, were not influenced by the presence of a group. Rarick, Widdop, and Broadhead suggested that modifications need to be made on existing tests for the EMR child, such as shortening the distance run. They stated that performance standards for normal children are not appropriate for EMR children, and that norms have to be set up for varying heights, weights, and ages in evaluating fitness tests. Jansma suggested that a definite diagnosis should not be made on the basis of a single test, and that the more seriously a student is handicapped, the more attention that must be paid to extraneous environmental variables.

CONCLUSIONS

The sit-up remains as the only available instrument to measure abdominal strength. The flexed arm hang can be modified with the proper equipment and will measure upper arm strength more reliably. The 12 minute and 2 mile runs are most reliable in measuring cardiovascular endurance. Substitutions for these runs include the step test, bicycle ergometer test, or several short dashes. However, modifications should be made when working with female subjects, younger children, and EMR children.

With the increased emphasis on skinfold assessment, educators need to be aware of the procedure for obtaining accurate measurements. When evaluating skinfolds, the less expensive skinfold calipers can be used, provided that the tester is experienced in using them. Educators also need to be aware of how to calculate ideal body weight because of our nation's obesity problem.

More trials need to be offered than those that are listed in various test manuals. However, the educator needs to be aware of the fatigue factor, especially in younger children. EMR children need many more test trials to achieve maximum performance than the normal child. All students will perform more consistently and reliably if they have had practice with the test items before the actual testing day.

Both normal and EMR children will perform better on a simple motor test if an audience is present because of the motivational factor. Both groups will perform much better with positive verbal feedback. However, a tangible reward seems to improve the performance level of the EMR child to a much greater extent.

The best score seems to be the most effective way to evaluate students when several trials are being performed if the threat of measurement error is small. When there is the chance of substantial measurement error and within-subject variation is small, the mean score seems to be the most accurate method of evaluation. Educators should be careful when using national norms for evaluation. However, norms are valuable in determining areas of need in the individual student.

Individual differences need to be taken into consideration when evaluating fitness. Boys and girls will vary as to the best components to use to test each aspect of fitness. Boys tend to show superiority in all tests at all ages, so achievement standards should be established separately. The EMR child will fall behind other students in performance levels, so achievement standards should also be modified for them. Age is an important factor when evaluating performance, but height and weight seem to have no influence.

Educators need to keep in mind the following key points when administering and evaluating physical fitness tests: 1) that they are familiar with the test and how to administer it; 2) that the test measures what it is intended to measure; 3) that more than one test is used to measure each component of fitness; 4) that they allow adequate rest between every item and follow the same order for each item; 5) that students are familiar with test items; 6) that motivation is provided; and 7) that proper scoring procedures are followed. The most prominent factor that was implied by most authors was the fact that each and every student is an individual and has specific areas of strengths and weaknesses. Physical educators need to consider all of the individual differences when compiling and evaluating fitness tests.

RECOMMENDATIONS

It is recommended that:

1. An acceptable substitute be found for the sit-up to measure abdominal strength.
2. Skinfold calipers be placed in all schools because of the increased awareness in body composition.
3. Fitness test manuals be revised to include more test trials in all tests where fatigue could not play a factor.
4. Physical educators allow practice days before the actual testing day for a fitness test.
5. More research be done on how to incorporate the EMR child into the regular classroom for fitness testing.
6. More research be done on school-age subjects rather than the college and adult population.
7. Most importantly, educators need to be aware of individual differences in the classroom and apply this knowledge when administering and evaluating fitness.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adams, F. A.; Linde, L. M.; and Miyake, H., "The Physical Working Capacity of Normal School Children in California," Pediatrics 28: 55-64, 1961. (9, 24, 25)

Baumgartner, Ted A., "Criterion Score for Multiple Trial Measures," Research Quarterly 45:193-197, 1974. (16, 24)

"Modified Pull-Up Test," Research Quarterly 49:80-84, March, 1978. (2, 9, 24)

Baumgartner, Ted A., and Jackson, Andrew S., "Measurement Schedules for Tests of Motor Performance," Research Quarterly 41:10-14, 1970. (14, 24)

Baumgartner, Ted A., and Zuidema, Marvin A., "Factor Analysis of Physical Fitness Tests," Research Quarterly 43:443-450, December, 1972. (17, 25)

Bowman, R. A., and Dunn, J. M., "Effects of Peer Presence on Psychomotor Measures with EMR Children," Exceptional Children 48:449-451, Fall, 1982. (3, 22, 26)

Cooper, Kenneth H., "Testing and Developing Cardiovascular Fitness," Journal of Physical Education Special Edition:130-144, March-April, 1972. (2, 11, 24)

Cooter, Rankin G., "Who Wants to be Normal?" Journal of Physical Education and Recreation 47:50, May, 1976. (16, 25)

Cotton, Doyice J., "A Modified Step Test for Group Cardiovascular Testing," Research Quarterly 42:91-95, March, 1971. (10, 24)

Disch, James; Frankiewicz, Ronald; and Jackson, Andrew, "Construct Validation of Distance Run Tests," Research Quarterly 46:169-176, May, 1975. (11, 24)

Falls, A. B., "Modern Concepts of Physical Fitness; AAHPERD Health-Related Physical Fitness Test," Journal of Physical Education and Recreation 51:25-27, April, 1980. (19, 26)

Fleischman, Edwin A., The Structure and Measurement of Physical Fitness, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1964, 207 pp. (2)

Franklin, Barry, "It's Almost Time for Another Invalid Appraisal of our Nation's Youth Fitness," Physical Educator 29:156, October, 1972. (19, 26)

Getchell, Bud, Physical Fitness: A Way of Life, John Wiley and Sons, Inc., New York, 1976 300 pp. (1, 5, 6, 7)

Hawkins, Jerald D., "An Analysis of Selected Skinfold Measuring Instruments," Journal of Physical Education, Recreation, and Dance 54: 25-27, January, 1983. (13, 24)

Hetherington, Ross, "Within-Subject Variation, Measurement, Error, and Selection of a Criterion Score," Research Quarterly 44:113-117, 1973. (15, 25)

Hunsicker, Paul, "Human Performance Factors," in Fitness, Health, and Work Capacity: International Standards for Assessment, pp. 349-361, edited by Leonard A. Larson, MacMillen Publishing Company, New York, 1974. (5)

Jansma, Paul, "Psychomotor Domain Tests for the Severely and Profoundly Handicapped," Journal of the Association for the Severely Handicapped 5:368-381, Winter, 1980. (23, 26)

Johnson, Barry L., and Nelson, Jack K., Practical Measurements for Evaluation in Physical Education, Burgess Publishing Company, Minneapolis, Minnesota, 1974, 438 pp. (5, 6)

Levy, Joseph, "Social Reinforcement and Knowledge of Results as Determinants of Motor Performance Among EMR Children," American Journal of Mental Deficiency 78:752-758, 1974. (21, 26)

Liba, Marie, "A Trend Test as a Preliminary to Reliability Estimation," Research Quarterly 33:245-248, 1962. (13, 24)

Lohman, Timothy G., and Pollock, Michael L., "Which Caliper? How Much Training?" Journal of Physical Education and Recreation 52:27-29, January, 1981. (12, 24)

Londeree, Ben R., "Strength Testing," Journal of Physical Education, Recreation, and Health 52:44-46, September, 1981. (17, 25)

Long, Boyde, "To Test Endurance," Athletic Journal 53:80-82, March, 1973. (10, 24)

Lowenbraun, Sheila, and Affleck, James Q., Teaching Mildly Handicapped Children in Regular Classes, Charles E. Merrill Publishing Co., Columbus, Ohio, 1976, 128 pp. (6)

Martens, Rainer, and Landers, Daniel, "Coaction Effects on a Muscular Endurance Task," Research Quarterly 40:733-737, 1969. (14, 25)

Mayhew, Jerry L., "Body Composition," Journal of Physical Education, Recreation, and Dance 52:38-40, September, 1981. (2, 12, 24)

Meyers, C. R., "In Quest of Sound Fitness Testing," Physical Educator 37:69-75, May, 1980. (2, 20, 26)

Montoye, H. J.; Frantz, M. E.; and Kozar, A. J., "The Value of Age, Heigt, and Weight in Establishing Standards of Fitness for Children," Journal of Sports Medicine and Physical Fitness 12: 174-179, 1972. (18, 25)

Nelson, J. K., and Dorociak, J. J., "Reducing Administration Time While Improving Reliability and Validity of Fitness Tests," Journal of Physical Education, Recreation, and Dance 53:63-65, January, 1982. (6, 7, 15, 25)

Olree, H., and others, "Evaluation of the AAHPER Youth Fitness Test," Journal of Sports Medicine and Physical Fitness 5:67-71, 1965. (1, 8, 26)

Plowman, Sharon A., and Falls, Harold B., "AAHPER Youth Fitness Test Revision," Journal of Physical Education and Recreation 49:22-24, November-December, 1978. (2)

Rarick, Lawrence G.; Widdop, James H.; and Broadhead, Geofrey D., "The Physical Fitness and Motor Performance of Educable Retarded Children," Exceptional Children 36:509-519, 1970. (22, 26)

Simri, Uriel, "Assessment Procedures for Human Performance," in Fitness, Health, and Work Capacity: International Standards for Assessment, pp. 362-377, edited by Leonard A. Larson, MacMillen Publishing Company, New York, 1974. (5, 6, 7, 20, 26)

Vincent, W. J., and Britten, S. D., "Evaluation of the Curl-Up: A Substitute for the Bent Knee Sit-Up," Journal of Physical Education and Recreation 51:74-75, February, 1980. (2, 8, 24)

Wagner, Harry Douglas, The Effects of Motivation and Repeated Trials on Physical Proficiency Test Performance of Educable Mentally Retarded Girls, Unpublished doctor's thesis, George Peabody College for Teachers, August, 1967, 91 pp., typed. (21, 26)

Webster's New Collegiate Dictionary, G. and C. Merriam Co., Springfield, Massachusetts, 1973, 1535 pp. (5, 6, 7)